

University of Groningen

Stable isotope quality assurance using the 'Calibrated IRMS' strategy

Meijer, Harro A. J.

Published in:
Isotopes in Environmental and Health Studies

DOI:
[10.1080/10256010902869113](https://doi.org/10.1080/10256010902869113)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2009

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Meijer, H. A. J. (2009). Stable isotope quality assurance using the 'Calibrated IRMS' strategy. *Isotopes in Environmental and Health Studies*, 45(2), 150-163. <https://doi.org/10.1080/10256010902869113>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Stable isotope quality assurance using the ‘Calibrated IRMS’ strategy[†]

Harro A.J. Meijer*

Centre for Isotope Research, University of Groningen, 9747 AG Groningen, The Netherlands

(Received 14 August 2008; final version received 13 November 2008)

Procedures in our laboratory have always been directed towards complete understanding of all processes involved and corrections needed etc., instead of relying fully on laboratory reference materials. This rather principal strategy (or attitude) is probably not optimal in the economic sense, and is not necessarily more accurate either. Still, it has proven to be very rewarding in its capability to detect caveats that go undiscovered in the standard way of measurement, but that do influence the accuracy or reliability of the measurement procedure. An additional benefit of our laboratory procedures is that it makes us capable of assisting the International Atomic Energy Agency (IAEA) with primary questions like mutual scale assignments and comparison of isotope ratios of the same isotope in different matrices (like $\delta^{18}\text{O}$ in water, carbonates and atmospheric CO_2), establishment of the ^{17}O – ^{18}O relation, and the replenishment of the calibration standards. Finally, for manual preparation systems with a low sample throughput (and thus only few reference materials analysed) it may well be the only way to produce reliable results.

Keywords: calibration; carbon-13; oxygen-18; IAEA; IRMS; isotope measurement methods and equipment; reference materials; water

1. Introduction

Ever since the introduction of the technique [1], the principle of dual inlet isotope ratio mass spectrometry (DI-IRMS) has been the fast comparison between sample and machine reference gas (MRG) under circumstances as identical as possible. In this fashion, the MRG was the basis of calibration. Regularly, isotope reference materials were prepared and measured with respect to this MRG, and these results made the MRG known on the international isotope scales. Since all preparation systems (such as for water and carbonates) were manual, there were only a few reference materials prepared on a daily basis, and one simply had to rely on the stability of the IRMS. The reference materials merely served as controls with which one could observe both clear errors in procedures and the long-term behaviour of the system.

In the course of time, however, when preparation systems became more and more automated, procedures started to change: usually one then had enough reference materials to calibrate the

*Email: h.a.j.meijer@rug.nl

[†]Updated paper: originally presented on the IAEA International Symposium “Quality Assurance for Analytical Methods in Isotope Hydrology” (August 2004, Vienna).

whole batch of samples just using their results on a daily basis without bothering anymore about the 'calibration' of the IRMS itself: the MRG became nothing more than just a Working Standard [2].

While this new strategy usually works well in the practical sense, the total measurement system suffers severely from loss of information. Since each batch of measurements is now self-contained, one need not observe the intermediate and long-term stability of the IRMS any longer, nor is the isotopic fractionation in the preparation system of importance anymore, as long as it is stable within one batch.

Keeping track of things like IRMS stability and stability of preparation systems, however, is crucial for the early discovery of potential problems. Furthermore, comparing the same isotope signal (e.g. $\delta^{18}\text{O}$) from different materials demands exact knowledge of the isotopic fractionation occurring in the different preparation systems.

At the CIO, therefore, we decided to follow a 'best of both worlds' strategy: We still keep our IRMS stable and well-calibrated, and we design and operate our preparation systems such that they either do not fractionate or do so in a stable and reproducible manner. On top of this, we use the abundance of reference material measurements from our preparation systems as a tool for final corrections while we strive for keeping these corrections insignificant.

This rather principal strategy (or attitude) is not optimal in the economic sense and is not necessarily more accurate either. Still, it has proven to be very rewarding in its capability to detect caveats that go undiscovered in the 'standard' way of measurement, but that do influence the accuracy or reliability of the measurement procedure. An additional benefit of our laboratory procedures is that they make us capable of assisting the International Atomic Energy Agency (IAEA) with primary questions like mutual scale assignments and comparison of isotope ratios of the same isotope in different matrices (like $\delta^{18}\text{O}$ in water, carbonates and atmospheric CO_2), establishment of the ^{17}O – ^{18}O relation, and the replenishment of calibration standards.

In the following, the author discusses the daily procedures followed at the CIO, illustrated with $\delta^{18}\text{O}$ measurements on water and carbonates.

2. Maintaining a calibrated IRMS

At the beginning of each measurement day, our DI-IRMS (a VG (now GVI) SIRA10) is loaded with MRG from a cylinder attached to it. The size of this cylinder is 0.5 l, and it is filled with CO_2 at typically 30 bar. At this pressure, all CO_2 is in the gas phase (condensation pressure \approx 58 bar at room temperature). This amount of MRG lasts for more than two years. The inlet procedure of the MRG is fixed: the MRG cylinder is mounted with two valves (Swagelok) with a small volume between them. First, this volume is evacuated, along with the reference side of the DI-IRMS itself. Then, MRG from the cylinder is loaded into the small volume while leaving the valve between this volume and the cylinder open for 30 s. Then, this valve is closed, and the one between the volume and DI system is opened, also for 30 s. Initially, when the MRG cylinder is at its initial pressure, the gas is expanded, not only into the bellows, but also into a larger part of the DI vacuum system. This volume is chosen such that the reference bellows are at about three quarters of their maximum volume while producing a flow of CO_2 through the capillary that produces the normal current at which measurements are being made (in our case 4 nA). In the course of time, the pressure of the MRG cylinder decreases. Its effect on the filling of the bellows is compensated by decreasing the total DI filling volume, until at some point only the bellows themselves are filled. The cylinder is put out of use as soon as the bellows cannot be filled to more than \approx 50 % of its volume. The pressure of the MRG cylinder has then decreased to somewhat below 10 bar.

After the loading of MRG, the IMRS is tested using five pure CO_2 samples, tapped from one of our two pure CO_2 local reference cylinders, named GS-27 and GS-28, respectively. The $\delta^{13}\text{C}$

and $\delta^{18}\text{O}$ values of these two are distinctly different, thereby giving an independent check (on a semi-daily basis) of the span of the isotope scales of the IRMS.

The two cylinders are normal high purity CO_2 cylinders containing liquid CO_2 . Therefore, samples from these cylinders will show long-term drifts. However, these drifts are very small: at most -0.01‰ per year for $\delta^{13}\text{C}$ (i.e. the liquid phase gets depleted in ^{13}C , an example of an anomalous isotope effect) and below $+0.02\text{‰}$ per year for $\delta^{18}\text{O}$. This means that they are perfectly suitable for daily tests. Even more, their behaviour over longer periods of time is valuable, too: their drifts are small and rather constant, so even observation over their entire lifetime is valuable.

Figure 1 shows results for GS-27 and GS-28. Table 1 gives their values, drift rates, and the standard deviation of a single measurement. All data are based upon analyses over the nearly seven year period shown in Figure 1.

The results of the measurement of these five samples of one of the cylinder gases are then tested on both their average value and their standard deviation. Test criteria are based upon the long-term experience with the gases, and with the IRMS. The criteria are also presented in Table 1. The total test result comes in the form of ‘malus points’ (a term we adopted from the insurance

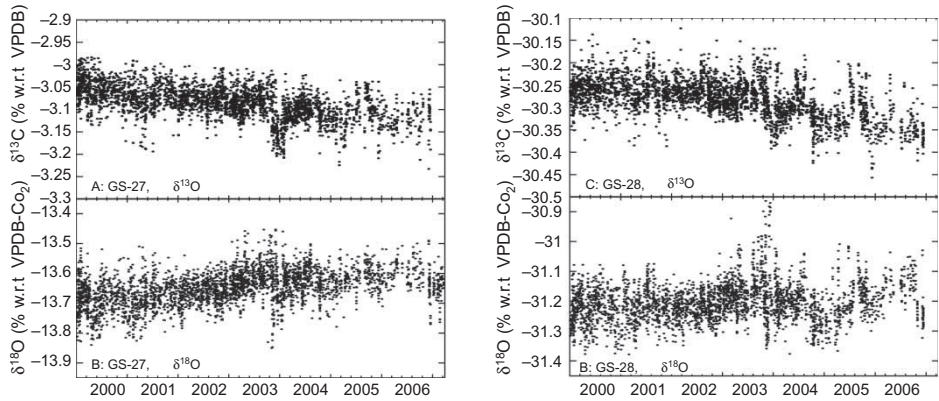


Figure 1. The results for GS-27 (A: $\delta^{13}\text{C}$, B: $\delta^{18}\text{O}$) and GS-28 (C: $\delta^{13}\text{C}$, D: $\delta^{18}\text{O}$) over a nearly seven year period. The small but steady drifts (negative for $\delta^{13}\text{C}$, positive for $\delta^{18}\text{O}$) can be clearly seen. The worth of daily testing can, e.g. be observed in December 2003, where we had problems with our IRMS. After recovery, the GS-27 and GS-28 show that the calibration situation of the IRMS has come back – within the error bars – to its original setting.

Table 1. Characteristics of the two reference gases in use for daily diagnostics of the IRMS.

					Test criteria (on average of five)	
		Average value 2000–2007, ‰ w.r.t. VPDB(– CO_2)	Drift rate, per meg/yr	Standard deviation (single mm, daily basis)	Deviation from average (‰) suspect–fail	σ -size (‰) suspect–fail
GS-27	$\delta^{13}\text{C}$	-3.083	-9.7 ± 0.3	0.030	0.02–0.03	0.03–0.04
	$\delta^{18}\text{O}$	-13.649	13.8 ± 0.5	0.047	0.05–0.08	0.08–0.12
GS-28	$\delta^{13}\text{C}$	-30.279	-11.1 ± 0.4	0.035	0.02–0.03	0.03–0.04
	$\delta^{18}\text{O}$	-31.209	4.6 ± 0.6	0.055	0.05–0.08	0.08–0.15

Note: The random error in the average values is below 0.001‰ , but the combined uncertainty is entirely due to the uncertainty in the scale maintenance. Drift rates are expressed in per meg ($=0.001\text{‰}$) per year. The drifts are (mostly) due to isotopic fractionation between the liquid and gaseous phase in the cylinders. Note the opposite sign of the drifts for $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$: whereas $\delta^{18}\text{O}$ of CO_2 shows the regular direction of isotopic fractionation, the $\delta^{13}\text{C}$ situation is anomalous: the heavier $^{13}\text{CO}_2$ molecule prefers the gas phase to the liquid phase. The test criteria apply to the results of the daily start-up test of the IRMS in which five samples of one of the two gases are measured. Their average is tested on deviation from the long-term average value. If the deviation is larger than the ‘suspect’ level, two ‘malus’ points are given, and three if it is larger than the ‘fail’ level. Similarly, for the standard deviation computed from the five measurements, the malus points are one and three, respectively. A total of three malus points declares the entire test ‘fail’, and appropriate measures have to be taken (see text).

world). Malus points are given when a quantity exceeds the inner limits around the expected value ('suspect'), and more when it exceeds the outer limits ('fail'). Exceeding the boundaries of the standard deviation is less severe than with the average value, since the uncertainty in the standard deviation determined with only five points is relatively large. The final test result can be 'OK', 'suspect' or 'fail'. In the case of 'fail', first a new set of five measurements is performed. If this set also fails, the inlet procedure of the reference gas is repeated. If the test again fails (and consistently in the same direction as the others), there is likely a problem with the IRMS and further diagnostics are necessary.

If the result of the measurements is 'OK' or 'suspect', measurements of samples may begin. Measurements performed are then – in principle – directly calibrated and normalised since we know our MRG – and also the reference cylinders GS-27 and GS-28 – with respect to VPDB- CO_2 (and VSMOW- CO_2). This knowledge has been gradually built up through frequent analysis of international reference and calibration materials, as well as local ones. These materials have been prepared using our sample preparation systems. This implies that our MRG is only well-calibrated if these preparation systems do not involve isotopic fractionation or if their isotopic fractionation is known and reproducible. During each day of measurements, these local and sometimes international reference materials are analysed along with samples from the respective preparation systems. These results are further indications of the calibration and stability of both the IRMS and the preparation systems.

At the end of the day, a set of three enriched CO_2 (typically +1000 ‰ for both $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$) samples is analysed. Their results serve to determine the so-called cross contamination [3] of the IRMS on a daily basis. Cross contamination is the apparent mixing of MRG and sample gas inside the IRMS, which gives rise to scale contraction: a true per mill difference is usually measured as less than a per mill. Cross contamination turned out to be the major source of inter-laboratory discrepancies, and it especially affects δ measurements relatively far away from the zero point of the δ scale. Using the 'absolute' ratios of a DI measurement of enriched CO_2 turns out to be a good way to determine cross contamination. Figure 2 shows long-term cross contamination results for our IRMS. The left-hand scale shows the amount of cross contamination η , and the right-hand scale shows the deviation of a true 40 ‰ difference that this cross contamination would cause. As can be clearly seen from Figure 2, the cross contamination is significant and even shows seasonal dependence: in summer it is higher than in winter. We can only speculate about the cause of this.

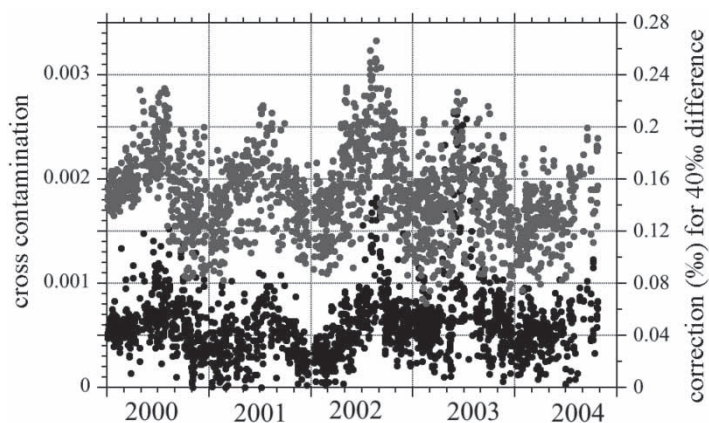


Figure 2. The long-term cross contamination δ (black, lower curve for mass 45, grey, upper curve, for mass 46) for our SIRA 10 machine. The right-hand scale gives the deviation of a true 40 ‰ difference between sample and MRG that this cross contamination would cause. Cross contamination is significant and even shows seasonal dependence: in summer it is higher than in winter. We suspect that this is due to the more humid atmosphere in our laboratories during summer.

While our laboratories are reasonably well-thermally stabilised, the relative humidity is larger in summer than in winter. We suspect tracers of water vapour entering the DI-IRMS vacuum system when changing the samples at the manifold inlet play the major role in this seasonality.

The final measurements of the day consist of three pure aliquots of CO₂ from our GS-27 or GS-28 cylinders. The first analysis of these after the enriched gases shows the sample-to-sample memory, which is about 0.02 % of the value difference for our system, and depends heavily on the pumping time of the inlet system between samples (120 s in our case). The next two give a final indication of stability of the system over the day.

Of course, our IRMS is subject to regular checks on items like electronic zero settings, balance of the capillaries ('zero enrichment'), the presence of small leaks etc. It turns out, however, that if problems of this kind start to occur, the IRMS normally will not pass our start-up test, and/or show drastic changes in cross contamination. This makes both the start-up test and gas measurements with CO₂ enriched in ¹³C and ¹⁸O very reliable diagnostic tools.

All measurement data, including the preparation details, are stored in a system of relational databases. This system not only stores the data, but also actively computes corrections (such as the cross contamination correction) and calibrates and normalises the data. An important feature of the system is that it allows for backward corrections in case these are necessary. Drifts, as well as sudden changes in calibration, are only well-interpretable after some time. Using the database system, a re-calibration or re-definition of a certain correction is easy to perform (and is documented).

3. Sample preparation and measurement

As explained in the previous chapter, our calibration strategy critically depends upon the performance of the preparation systems. Therefore, we have designed our preparation lines to resemble the original guidelines as closely as possible [4,5].

3.1. Automatic oxygen from water (AOW) preparation system

In the case of water preparation, the VSMOW-SLAP $\delta^{18}\text{O}$ scale has been defined based on the isotopic equilibration of water with CO₂ (first described in [6]):



The isotopic equilibrium depends critically on the temperature. If one exclusively measures water samples, the choice of temperature does not matter as long as it is kept constant. If one also measures other oxygen-containing compounds (such as carbonates) and wants to compare the isotopic composition of the CO₂ evolved from these different compounds, the absolute value of the temperature is important.

Coplen *et al.* [7] compare CO₂ evolved from VSMOW at exactly 25 °C ('VSMOW-CO₂') with CO₂ evolved from NBS 19 also at 25 °C (see further). Their results are now an essential part of the recommended set of fractionation factors of the water and carbonate (see below) preparation schemes [4,5,8], all at 25 °C. This has led to the situation in which the $\delta^{18}\text{O}$ of CO₂ in equilibrium with VSMOW at exactly 25 °C determines the zero-point of the so-called VSMOW-CO₂ scale. It is this scale that we want to reproduce.

Commercial systems for automatic $\delta^{18}\text{O}$ measurement of water are available as an add-on to IMRS machines. They usually work at elevated temperatures, and they evacuate the water vessel by pumping through a capillary. This gives rise to evaporation of some of the water, and thus to isotopic fractionation. As long as circumstances are identical for all samples, and there are

enough reference materials in a batch, results of such systems can be very precise. However, these systems are not designed to exactly reproduce the VSMOW-CO₂ scale, and thus do not fit our calibration strategy.

Therefore, we designed our own automatic preparation system in which we take many precautions to prevent any spurious isotopic fractionation effects, and we correct effectively for all known isotopic fractionations. The system's capacity is 80 samples, arranged in five subsets of 16 reaction vessels each. We use typically 0.6 ml water and 0.3 mmol of CO₂ gas. The water-CO₂ equilibrium is established statically in at least 24 h, at exactly 25.00 °C (± 0.02 °C), so that the CO₂ after establishment of the equilibrium should be exactly on the VSMOW-CO₂ scale.

The final result of the equilibrium process is slightly influenced by the added CO₂, which is initially not in isotopic equilibrium with the water. We determine this so-called 'water correction' by measuring the quantity of CO₂ upon extraction. To keep this correction small, we have two CO₂ cylinders available with distinctly different isotopic composition. One is for the high end of the natural range for $\delta^{18}\text{O}$ (−25 to 0 ‰), and the other is for the low end, e.g. ice cores from Greenland and Antarctica (< −25 ‰). We have tested the correctness and precision of our determination of the water correction by running all local reference waters with both equilibrium CO₂'s, as well as by performing the equilibration with isotopically enriched CO₂ ($\delta^{18}\text{O}$ of approximately +1000 ‰).

After putting water samples into the vessels, we evacuate these vessels. To prevent isotopic fractionation through evaporation effects, we freeze the water sample in dry ice before evacuation. We also remove the gases originally dissolved in the water: after evacuation we close the vessels, thaw the water (now under vacuum conditions so that most of the dissolved gases get transferred to the head space), refreeze it again, pump again and then add the CO₂ to the still frozen water. Then the vessels are put (statically) in the 25.00 °C water bath for at least 24 h.

For analysis of the CO₂, the system with the water vessels is coupled to the DI-IRMS, and measurements proceed automatically. One after the other, the vessels are opened and the CO₂ is led, in a two-step process through a cold trap at ≈ -55 °C to freeze out water vapour, into a liquid N₂ cooled trap in which all CO₂ is collected. This CO₂ is then led after thawing and measurement of its quantity into the DI-IRMS.

3.2. Carbonates preparation system

As our laboratory is not involved in carbonate measurement programs, and we mainly use carbonates for calibration purposes, our 'single batch' carbonate preparation line, containing 10 sample positions, is manual and off-line. It is designed in line with the original paper of McCrea [9]. We use typically 20 mg of carbonate material per sample. It is put into a special reaction vessel together with 4 ml of ≥ 100 % (s.g. = 1.93) phosphoric acid in a side-arm, not yet in contact. The vessel is evacuated and then put into a water bath at 25.00 (± 0.05) °C. After about 1 h, the reaction is started by tilting the reaction vessel. After a reaction time of at least 18 h, the evolved CO₂ is extracted through a vacuum line – through a dry ice cold trap to collect water vapour – and collected in a liquid nitrogen cold trap. From this cold trap, the CO₂ is frozen over in a small sample tube, which is then connected to the sample manifold of the IRMS for isotopic analysis.

4. Performance of the systems

4.1. Water samples

The calibration of the combination AOW preparation system and DI-IRMS has to proceed using the primary calibration material VSMOW. In practice, however, we maintain a set of seven local reference waters that cover the whole natural range from ocean water down to Antarctic precipitation.

Table 2. Description of the seven local reference waters in use by the CIO.

Name	Water type	$\delta^{18}\text{O}$ (‰ w.r.t. VSMOW-SLAP)
GS-49	Distilled seawater	0.39
GS-48	Distilled tap water	−6.52
GS-22	Alpine snow	−15.29
GS-42	Greenland ice	−24.62
GS-50	Ice water mix	−35.06
GS-46	Antarctic ice	−43.69
GS-47	Antarctic ice	−50.69

These waters are stored in closed containers with dispensers. Thanks to frequent calibration runs in our laboratory over the course of the years, several ring tests in which we participated (Inter-laboratory tests by the IAEA isotope hydrology section [10–12], ring test within the European Project on Ice Coring in Antarctica) and to a bilateral intercomparison with the IAEA isotope hydrology laboratory, the δ -values of these waters are very well calibrated. Table 2 lists the reference materials.

In the normal measurement routine, each subset of 16 samples contains four of these local reference materials: two samples of a reference in the high end, and two at the low end of the expected range of the unknowns. Regularly, an extra reference sample is added with its composition as far away from the sample range as possible.

The reproducibility within a single batch with all subsets taken together is better than 0.04 ‰, regardless of the δ -value of the reference material. This is the same performance as our DI-IRMS shows for single flasks filled with pure CO_2 (or even slightly better, cf. Table 1). The long-term combined uncertainty is necessarily worse than that: ± 0.06 ‰ over the whole range, however, with occasional outliers (see further).

This performance is reached relying on our ‘calibrated IRMS’ strategy so that no further batch-wise normalisation has been performed. Figure 3 shows histograms of results for four of our reference waters over a period of six months.

While the distribution of three of them is reasonably normal, the water most depleted in ^2H and ^{18}O shows asymmetric ‘tailing’. The cause for this is not entirely clear, but we suspect that sample-to-sample memory effects play a role. Their relatively large influence would then be caused by the addition of widely differing reference materials to a set of samples as described above. This is observed when the reference GS-47, which is depleted in ^2H and ^{18}O , is being used for that purpose. In this case, the ability of being able to normalise the batch using co-measured reference materials comes at the price of memory effects becoming significant. A slight advantage, on the other hand, is that we can now observe – and possibly eliminate – these memory effects. Despite our design and maintenance efforts, sometimes whole batches deviate significantly and consistently. The cause of such a deviating day is usually unknown. Start-up and closing samples on the IRMS that day show no anomaly. If all reference waters in such a deviating batch consequently show the same deviation, we use their values to perform a batch normalisation in our database system.

The most remarkable result of our ‘calibrated IRMS’ strategy is the relation between the δ -values measured for our local reference waters and their normalised ones. According to our strategy, there should be no need for scale normalisation since we have taken all known precautions to make sure that one per mil difference in our measurements is as accurate as possible. However, our measured values for our local reference waters do not correspond with their normalised assigned values. Figure 4 shows the deviation between measured and assigned normalised values of our reference (as well as of VSMOW) as a function of the assigned value. The error bars are our best estimate for the combined uncertainty in the mean.

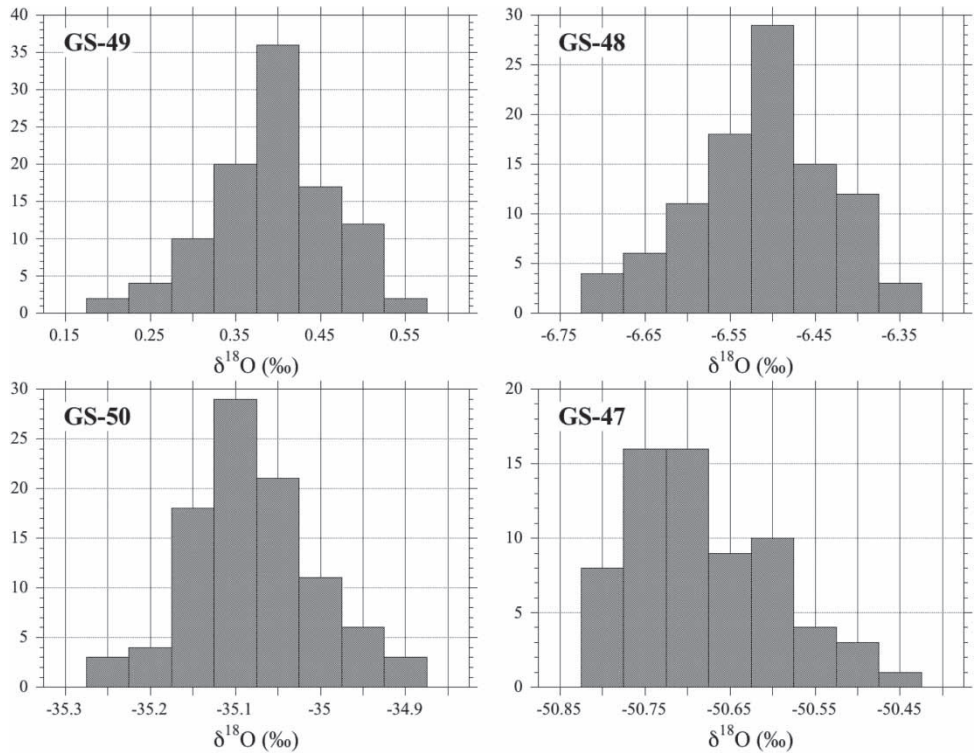


Figure 3. Histograms of results for four of our reference waters over a period of six months. This performance is reached relying on our ‘calibrated IRMS’ strategy so that no further batch-wise normalisation has been performed. Three of the waters show a reasonably normal distribution, having a standard deviation of 0.06 ‰. The water most depleted in ^2H and ^{18}O (GS-47), however, shows asymmetric ‘tailing’. We suspect that sample-to-sample memory effects play a role.

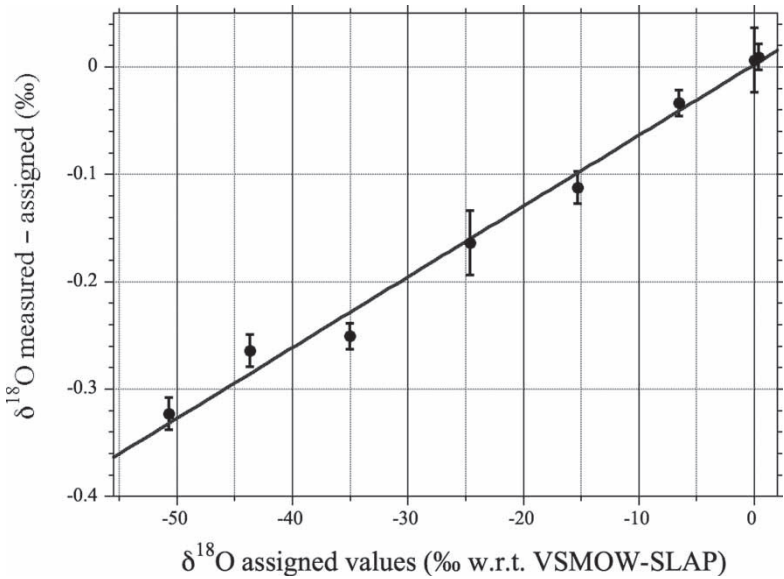


Figure 4. The deviation between measured and assigned normalised values of our reference waters (as well as of VSMOW) as a function of the assigned value. There is a clear trend in the deviations from zero difference at the position of VSMOW (which is the direct effect of the calibration procedure) to -0.36 ‰ at SLAP. The most likely explanation of this phenomenon is that the assigned value to SLAP of -55.5 ‰ is too positive.

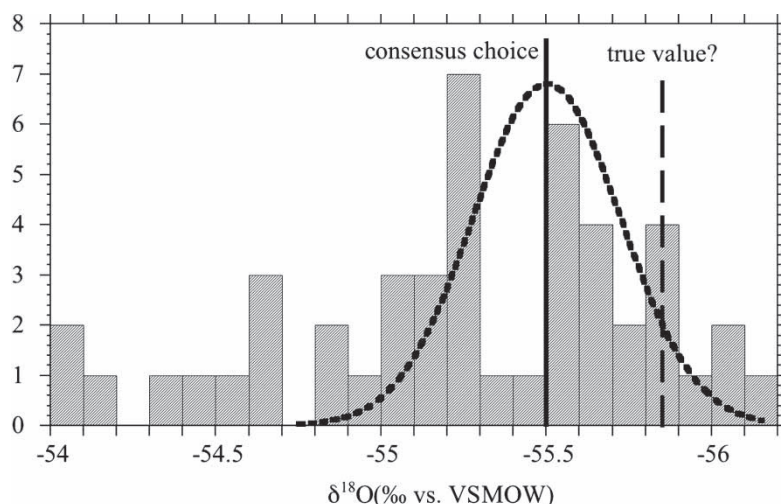


Figure 5. Results of the original determination of SLAP vs. V-SMOW (measurements quoted in [5]) on the basis of which the value of -55.5 ‰ was chosen 'in consensus', (according to the assumed normal distribution sketch in the figure). There is a large spread of results, caused by many effects that researchers were not all aware of at that time, most notably cross contamination. Most of these effects cause a systematic deviation towards less negative numbers. According to our findings, a value of -55.85 ‰ for SLAP would be more accurate.

There is a clear trend in the deviations, accurately described by the shown linear fit, from zero difference at the position of VSMOW (which is the direct effect of the calibration procedure) to -0.36 ‰ at SLAP.

The most likely explanation of this phenomenon is that the assigned value to SLAP of -55.5 ‰ is too positive. Figure 5 shows the original exercise [5,13,14] on the basis of which this value of -55.5 was chosen 'in consensus'. There is a large spread of results, caused by many effects that researchers were not all aware of at that time, most notably, cross contamination. Most of the effects cause a systematic deviation towards less negative numbers. According to our findings, a value of -55.85 ‰ would be more accurate. This finding is supported by the results of a recent selective ring test, involving cross-contamination correction, initiated by the CIO and organised by NIST [15,16]. Their result for SLAP is even more negative ('between -55.7 and -56.2 ‰').

Of course, we report $\delta^{18}\text{O}$ measurements of water on the VSMOW–SLAP scale. The substantial benefit of this normalisation procedure is that when it is used in laboratories worldwide, isotopic analysis of the same sample by two or more laboratories should yield the same value within analytical uncertainty. Ironically, however, this convention forces us to contract the scale of our measurements.

Since local reference waters are analysed almost every day, the AOW system provides us with a high quality extra record to maintain and check our calibrated ^{18}O scale.

4.2. Carbonate samples

The calibration of the VPDB– CO_2 scale is based on analysis of NBS 19 samples. In practice, we realise and maintain this calibration using several local carbonate reference materials, a few of which are in use over decades. Thanks to the large number of comparisons between these local reference materials and NBS 19 and international reference materials (such as NBS 18, IAEA-CO-1 etc.), we have come to know our own reference materials very well indeed.

The single sample standard deviations are ± 0.04 ‰ and ± 0.10 ‰, for $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$, respectively. For $\delta^{13}\text{C}$, there are usually no systematic problems, and calibration exercises usually

reproduce over the years at the ± 0.01 ‰ level. Correction for cross contamination, however, is essential. Unlike the VSMOW–SLAP scale, which has been established relatively early in the history of isotope measurements, a second reference point on the $\delta^{18}\text{O}$ scale has not been defined for decades. Many laboratories, though, have used one of the international reference materials depleted in ^{13}C , LSVEC, or IAEA-CO-9, for this purpose [17], and it has been recommended that laboratories state that the isotopic composition of an appropriate internationally distributed reference material had it been analysed with the unknowns [18].

Still, no consensus values existed for these reference materials until 2006. Their values had been re-established in the afore-mentioned ring test [16], in which cross-contamination effects have been accounted for. Our results for these light materials agreed well with the results of the test (the CIO was lab #6 in the test).

Only recently, the material L-SVEC (a lithium carbonate) has been officially prescribed and recommended as second calibration point to normalise the scale [19], as the result of an intensive collaborative effort. Its recommended value, -46.6 ‰, is only slightly more negative than the result in the earlier ring-test mentioned above, which was -46.48 ‰ [16].

For $\delta^{18}\text{O}$, the carbonate reaction is much more critical since both water and CO_2 are formed and (kinetic) isotopic fractionation is involved. Furthermore, there is the danger that the $\delta^{18}\text{O}$ values of carbonates in stock are influenced by oxygen exchange with water vapour over a period of years. In spite of these two facts, our representation of the VPDB- CO_2 scale shows good stability over the years. Like in the $\delta^{13}\text{C}$ case, the cross-contamination correction for $\delta^{18}\text{O}$ (about twice as large for $\delta^{18}\text{O}$ as for $\delta^{13}\text{C}$ for our IRMS) is very important.

As said, our manual system produces only 10 samples per batch. For calibration purposes this is sufficient. If such a manual system is used for unknown samples, however, not many samples can be produced per day and only a few local reference materials can be analysed. Each batch would typically contain only two reference materials. In such a case, batch-wise normalisation involves large errors and large risks (if one of the two reference materials were in error, the whole batch would be assigned an erroneous calibration). For these circumstances, our strategy of a calibrated IRMS has large advantages.

5. Comparison of ^{18}O scales: VPDB- CO_2 and VSMOW(-SLAP)- CO_2

Typically twice per year (but dependent on the trends we observe in all our reference materials over time), we perform an extensive ‘calibration exercise’, in which calibration materials (carbonates and waters) and reference materials (carbonates, waters, and pure CO_2) are prepared on our preparation systems and analysed in sequence on our IRMS. The purpose of such an exercise is to check and re-establish the existing calibration of the DI-IRMS of which the MRG is the crucial element.

The spin-off of such an exercise is the difference between the two ^{18}O scales VPDB- CO_2 and VSMOW- CO_2 . Our best effort ‘calibrated IRMS’ strategy includes preparation systems that work as exactly as possible according to the original definitions of the scales. Therefore, comparison of the water and carbonate calibration materials (supported by both international and local reference materials) should directly yield the difference between these scales, or in other words: we should directly get the $\delta^{18}\text{O}$ value of VSMOW- CO_2 on the VPDB- CO_2 scale and vice versa.

Seen from another perspective, the difference we obtain between our best effort to represent both scales, and its stability over the years, acts as a quality mark for our efforts. The scale difference has been determined by Coplen *et al.* [7] to be -0.26 ‰ (VSMOW- CO_2 w.r.t. PDB- CO_2). This value, adjusted by -0.01 ‰ to account for the subtle difference between PDB and VPDB measured by Coplen *et al.* [7], has been included in the ‘consensus’ scheme in which both scales are represented with their consensus isotopic fractionation factors for the preparation,

as well as the differences between the different materials involved (the virtual carbonate VPDB, the carbonate NBS 19, the CO₂ evolved from them through the phosphoric acid reaction, the reference water VSMOW and the CO₂ in isotopic equilibrium with it) [4,8,14]. Of course, like the consensus value for SLAP, the recommended value of the VSMOW-CO₂/VPDB-CO₂ scale difference bears uncertainty as well.

Figure 6 shows the difference between the two $\delta^{18}\text{O}$ scales, as observed in our system. The waters (in black, the same points and curve as in Figure 4) are expressed with respect to VSMOW-CO₂, the carbonates (in grey) with respect to VPDB-CO₂. However, the latter scale has not been independently determined, but is calculated based on the established VSMOW-CO₂ scale using the consensus difference of 0.27 ‰. The y-axis shows the difference between the (non-normalised) measurements and the assigned values. The fit curve through the water points is in fact the normalisation procedure, fixing the value of -55.5 ‰ for SLAP, and proportionally less negative values for our references. The fit curve through the carbonates is what a normalisation procedure would produce that would rely on the measurements shown: the calibration material NBS 19, the international reference material NBS 18, and four of our local references. The water results yield zero at their calibration value of 0 ‰. This is no surprise, as the scale is produced using these waters. Ideally, the carbonate results should yield zero deviation at their calibration value of -2.20 ‰. Our laboratory results, however, show a slight deviation from zero for the carbonates, of between 0.04 and 0.05 ‰. This implies that our result for the VSMOW-CO₂/VPDB-CO₂ difference would be about 0.32 ‰. It is hard to estimate the combined uncertainty in this result. In spite of our care with the two preparation methods, however, systematic deviations of this size might well be introduced by either one of the preparation lines.

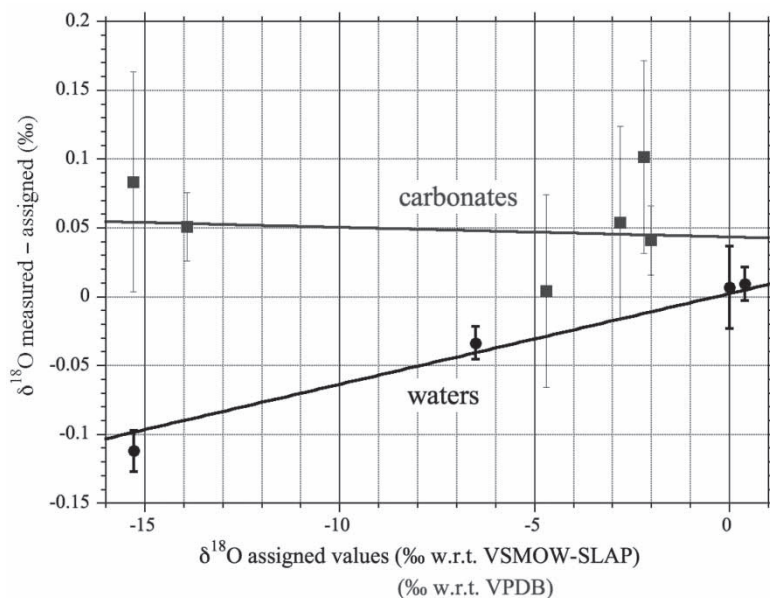


Figure 6. The difference between the two $\delta^{18}\text{O}$ scales, as observed in our system. The waters (in black, the same points, and curve as in Figure 4) are expressed with respect to VSMOW-CO₂, the carbonates (in grey) with respect to scale calculated based on VSMOW and the consensus difference of 0.27 ‰. This should ideally reproduce the VPDB-CO₂ scale. The y-axis shows the difference between the (non-normalised) measurements and the assigned values. The water results yield zero at their calibration value of 0 ‰, as they should. Ideally, the carbonate results should yield zero deviation at their calibration value of -2.20 ‰. Our laboratory results, however, show a slight deviation from zero for the carbonates, of between 0.04 and 0.05 ‰. This implies that our result for the VSMOW-CO₂/VPDB-CO₂ difference would be about 0.32 ‰. (For further explanation see text).

Before we had our present automatic water preparation system, we used a manual one, which had a similar design. Yet, using that system the difference between the VPDB- and VSMOW-CO₂ scales tended to be slightly smaller than 0.27 ‰. The NIST ‘cross-contamination’ ring test [16] produces a value for the difference between the scales of 0.17 ± 0.26 ‰, where the error bar (the combined uncertainty) shows the relatively large spread between the eight expert participants.

Regarding these results, there is no reason to doubt the value of 0.27 ‰ between the scales, based on [7] and recommended since 1983. It was not our intention to re-determine this number (then several measurement sessions with multiple samples of VSMOW and NBS 19 would be more adequate), but rather to check our representation of the two $\delta^{18}\text{O}$ scales with respect to each other. In that sense, we are satisfied with the result. In practice, we report results for waters using the normalisation fit shown (see Figure 4), and for carbonates, the normalisation shown in Figure 6. The difference between the normalisation lines (Figure 6) is then a direct measure of the reliability of our respective scale representation. While this does not matter much for the analysis of carbonates or waters themselves (only reproducibility of the preparation is crucial then), it does matter when analysing other materials, most notoriously atmospheric CO₂.

Isotopic measurements on atmospheric CO₂ require the utmost precision, most of all for $\delta^{13}\text{C}$, but also for $\delta^{18}\text{O}$. For $\delta^{13}\text{C}$, the VPDB scale is used. As the reaction from carbonate to CO₂ involves no isotopic fractionation for carbon, this scale is quite straightforward to realise (although the required 0.01 ‰ level remains a continuous challenge). For $\delta^{18}\text{O}$, on the other hand, the ‘VPDB-CO₂’ scales maintained at the various laboratories worldwide involved in atmospheric CO₂ were different by as much as 0.5 ‰ in the 1990s. These differences were clearly shown by pure CO₂ isotopic reference gases that the CIO produced [20]. Intercomparison between just a few groups [21] showed that calibration problems for $\delta^{18}\text{O}$ were larger by one order of magnitude than for $\delta^{13}\text{C}$. The reason for this difference is the isotopic fractionation involved in the $\delta^{18}\text{O}$ scale definitions. Most laboratories involved in isotope measurements of atmospheric CO₂ are not involved in carbonate measurements, water measurements, or perhaps either of them. Even if they are, they usually use commercial equipment designed for high-precision reproducibility, but not for scale representation according to the original definitions.

In order to perform well-calibrated $\delta^{18}\text{O}$ measurements, especially on substances other than carbonates or water, we regard it as crucial to try to represent the two scales independently in the laboratory and to use the scale difference as a final check on quality.

6. Conclusions and final remarks

Our ‘calibrated IRMS’ strategy is in the first place very satisfactory from a physics’ point of view. Instead of performing daily ‘batch-wise’ corrections of varying size, for which the reasons are unknown, one tries to operate all parts of the metrology system as optimally as possible. If deviations still exist, or occasionally happen, one tries to find the causes for these deviations, and eliminate them. Only as a last resort, and as rarely as possible, one has to apply final normalisation factors.

At the CIO, we have found this strategy very robust over the years and intrinsically more reliable. This applies even more strongly for $\delta^{18}\text{O}$ than for $\delta^{13}\text{C}$ since for $\delta^{18}\text{O}$ realisation of both scales in use involves isotope-fractionation processes. Realizing both scales at the same time has proven to be a valuable check on the realisation of both of them. An additional advantage is that primary calibration and normalisation questions, as illustrated above, can be dealt with.

One disadvantage of the strategy is that it is more cumbersome to maintain. Therefore, it is not likely to be a strategy that isotope measurement ‘production’ laboratories are inclined to follow. Rather, they would rely on calibration and reference materials prepared for them by institutions

like IAEA and NIST. However, these institutions still need assistance of a significant number of (university) laboratories that do follow a more principled calibration strategy, like the one sketched in this work. One already observes the tendency of 'special purpose' ring tests (e.g. for the re-establishment of assigned values to certain reference materials) to be organised among a selected group of invited laboratories only. The cross-contamination ring test [16] is a nice example of this.

We have also tried to apply the calibrated IRMS strategy to continuous flow IRMS. In principle, the strategy can be applied equally well: one can regularly calibrate the reference gas, apply blank and mass dependence corrections as well as possible etc. Still, final reference material-based batch corrections remain necessary, and these can be relatively large (several tenths of a per mil), and quite variable. There are too many processes involved that can lead to a variable and unknown isotopic fractionation, such as the conditions of combustion and reduction reactors, (open) splits in capillaries, etc. These effects have made our calibrated 'CF-IRMS' strategy until now not very rewarding. Nevertheless, monitoring the batch corrections from batch to batch gives quite sensitive extra information on the status of the system as a whole.

Cross contamination, as it occurs in DI instruments, obviously does not exist in CF. Taking the necessary precautions, such as adequate blank correction, and using samples of equal C content, a measured per mil difference can be quite accurate, as the collaborative effort described in [19] shows.

Acknowledgements

The author would like to thank the technical team of the CIO for their continuous high-quality work throughout the years, especially H.A. Been, J. Schut, and W. Jonker for their development work on the AOW system, Mrs. J.J. Spriensma for water and carbonate sample preparation and H.G. Jansen for maintenance of and measurements on the IRMS. Our QA/QC manager Mrs. A.T. Aerts-Bijma's help with the article is gratefully acknowledged. Thanks to Dr. Tyler B. Coplen for his valuable comments on the manuscript.

References

- [1] C.R. McKinney, J.M. McCrea, S. Epstein, H.A. Allen, and H.C. Urey, Improvements in Mass Spectrometers for the Measurement of Small Differences in Isotope Abundance Ratios, *Rev. Sci. Instrum.* **21**, 724 (1950).
- [2] R.A. Werner and W.A. Brand, Referencing Strategies and Techniques in Stable Isotope Ratio Analysis, *Rapid. Commun. Mass Spectrom.* **15**, 501 (2001).
- [3] H.A.J. Meijer, R.E.M. Neubert, and G.H. Visser, Cross Contamination in Dual Inlet Isotope Ratio Mass Spectrometers, *Int. J. Mass Spectrom.* **198**, 45 (2000).
- [4] R. Gonfiantini, W. Stichler, and K. Rozanski, Standards and Intercomparison Materials Distributed by the International Atomic Energy Agency for Stable Isotope Measurements, in *Reference and intercomparison materials for stable isotopes of light elements*, IAEA-TecDoc 825, 1995. <www-pub.iaea.org/MTCD/publications/>.
- [5] R. Gonfiantini, Advisory Group Meeting on Stable Isotope Reference Samples for Geochemical and Hydrological Investigations, IAEA Report to the Director General, 1984.
- [6] S. Epstein and T. Mayeda, Variation of O^{18} Content of Water from Natural Sources, *Geochim. et Cosmochim. Acta* **4**, 213 (1953).
- [7] T.B. Coplen, C. Kendall, and J. Hopple Comparison of Stable Isotope Reference Samples, *Nature* **302**, 236 (1983).
- [8] G. Hut, Consultants' Group Meeting on Stable Isotope Reference Samples for Geochemical and Hydrological Investigations, IAEA Report to the Director General, 1987.
- [9] J.M. McCrea, On the Isotopic Chemistry of Carbonates and a Paleotemperature Scale, *J. Chem. Phys.* **18**, 849 (1950).
- [10] L. Araguás Araguás and K. Rozanski, Interlaboratory Comparison for Deuterium and Oxygen-18 Analysis of Precipitation Samples, IAEA Report, Vienna, 1995.
- [11] J. Lippmann, M. Gröning, and K. Rozanski, 2nd Interlaboratory Comparison for Deuterium and Oxygen-18 Analysis of Water Samples, IAEA Report, Vienna, 1999. <http://www.iaea.org/programmes/rial/pci/isotopehydrology/i_c_c.shtml>.
- [12] M. Gröning, S. Graeber, S.K. Bhattacharya, M. van Duren, and L. Andreescu, 3rd Interlaboratory Comparison Exercise for δD and $\delta^{18}O$ Analysis of Water Samples (WICO2002-Part I), IAEA Report, Vienna, 2003. <http://www.iaea.org/programmes/rial/pci/isotopehydrology/i_c_c.shtml>.
- [13] R. Gonfiantini, Consultants' Meeting on Stable Isotope Reference Samples for Geochemical and Hydrological Investigations, IAEA Report to the Director General, 1977.
- [14] R. Gonfiantini, Standards for Stable Isotope Measurements in Natural Compounds, *Nature* **271**, 534 (1978).

- [15] H.A.J. Meijer and D.B. Klinedinst, 8th IAEA Advisory Group Meeting on Future Trends in Stable Isotope Reference Materials and Laboratory Quality Assurance, Vienna, Austria, 2000. <<http://www.iaea.org/programmes/rial/pci/isotopehydrology/meetings.shtml>>.
- [16] R.M. Verkouteren and D.B. Klinedinst, Value Assignment and Uncertainty Estimation of Selected Light Stable Isotope Reference Materials: RMs 8543–8545, RMs 8562–8564, and RM 8566 NIST Special Publication (SP 260–149), 2003. <http://patapsco.nist.gov/srmcatalog/sp_publications/documents/sp260-149.pdf>.
- [17] W. Stichler, Interlaboratory Comparison of New Materials for Carbon and Oxygen Isotope Ratio Measurements, in *Reference and intercomparison materials for stable isotopes of light elements*, IAEA-TecDoc 825, 1995. <www-pub.iaea.org/MTCD/publications/>.
- [18] T.B. Coplen, New Guidelines for the Reporting of Stable Hydrogen, Carbon, and Oxygen Isotope Ratio Data, *Geochem. Cosmoch. Acta* **60**, 3359 (1996).
- [19] T.B. Coplen, W.A. Brand, M. Gehre, M. Gröning, H.A.J. Meijer, B. Toman, and R.M. Verkouteren, New Guidelines for $\delta^{13}\text{C}$ measurements, *Anal. Chem.* **78**, 2439 (2006).
- [20] H.A.J. Meijer, The Isotopic Composition of the Groningen GS-19 and GS-20 Pure CO_2 Standards, in *Reference and Intercomparison Materials for Stable Isotopes of Light Elements*, IAEA-TecDoc 825, 1995. <www-pub.iaea.org/MTCD/publications/>.
- [21] H.A.J. Meijer, Reference Materials and Intercomparison studies, Chapter SI-04 in CIO Scientific Report 1995–1997, 1998. Table on GS-19 and GS-20 reproduced in [2].